IDENTIFICATION AND CONVEXITY IN OPTIMIZING CONTROL

Bjarne A. Foss [**]

[*] *Department of Engineering Cybernetics, Norwegian University of Science and Technology, N-7034 Trondheim, Norway. Email: baf@itk.unit.no*

Tor A. Johansen [**]

[**] *SINTEF Electronics and Cybernetics, Automatic Control Department, N-7034 Trondheim, Norway. Email: Tor.A.Johansen@ecy.sintef.no.*

Abstract. The interplay between identification of a predictor and the convexity of an associated optimization-based control problem is studied. The reason for the appearance of nonconvexity is investigated and a method to favour convexity of the optimization-based control problem in the identification problem is suggested. Further, a parametrization which makes the identification problem tractable is proposed. Simulation results suggest a merit to method.

Keywords. Optimization, nonconvex problems, identification, nonlinear systems, model predictive control

1. INTRODUCTION

The combined use of dynamic models and optimization offers a concept in which process knowledge can be linked to operational goals formulated by some optimization criterion. This concept has seen widespread use. One such area is process control and the use of model predictive control (MPC). MPC refers to a class of algorithms where an optimization problem is solved repetitively, at every new time-instant, see for example Lee (1996). The interaction between control and optimization is discussed in an illuminating way in Mayne (1995). In this paper the divide between convex and nonconvex problems is emphasized. In particular problems arise by using a nonlinear predictor as opposed to a linear predictor since the optimization problem in the former cases in general becomes nonconvex. This latter observation forms the departure point for this study.

The paper is structured as follows: First, a problem definition is provided. Thereafter, we investigate convexity in nonlinear optimization-based control. In particular, we discuss design tradeoffs between identifying a predictor, and convexity of an associated optimization problem. Section 4 highlights some of the findings by a simulation example, while a discussion and some conclusions ends the paper.

2. PROBLEM DEFINITION

The optimality criterion is defined in discrete-time

$$\phi_k(z_k) = \phi_k(\pi_k, \chi_k) = \sum_{i \in I_k} l_i(x_{i+1}, u_i) \quad (1)$$

where

$$\begin{aligned}
\pi_k &= \{u_k^T, \ldots, u_{k+N-1}^T\}^T, \qquad u_i \in U \subseteq R^{m_u} \\
\Pi &= U \times \ldots \times U \\
\chi_k &= \{x_{k+1}^T, \ldots, x_N^T\}^T, \qquad x_i \in X \subseteq R^{m_x} \\
\mathcal{X} &= X \times \ldots \times X \\
I_k &= \{k, \ldots, k+N-1\}
\end{aligned}$$

u_i denotes the control input of a system. Since we are assuming a time-discrete model the control input u_i is constant during the time span $[i, i+1>$. x_i denotes the system state at time i.

Above, the sets Π and $\mathcal{X}$ (or U and X) invoke constraints on π_k and χ_k. The states and control inputs are, however, not independent, since the state sequence χ_k in a dynamic system is uniquely defined by the initial value, ie. x_k, and the control input sequence π_k. We define a state-space model, valid on at least I_k, and a predictor by

$$h_i(x_{i+1}, x_i, u_i) = x_{i+1} - f_i(x_i, u_i) = 0 \quad (2)$$

$$\begin{aligned}
\chi_k &= F_k(\pi_k, x_k) \quad (3) \\
&= \{F_{k1}(\pi_k, x_k)^T, \ldots, F_{kN}(\pi_k, x_k)^T\}^T
\end{aligned}$$

which means that F_k performs the mapping $F_k : \Pi \times X \rightarrow \mathcal{X}$. The predictor can be realized by multiple use of a state-space model.

The minimization problem can be stated as

$$\pi_k^o = arg\ min_{\pi_k \in \Pi}\ \phi_k(\pi_k, \chi_k) \quad (4)$$

subject to

$$\begin{aligned}
&\chi_k = F_k(\pi_k, x_k) \in \mathcal{X} \\
&x_k - \text{given}
\end{aligned}$$

For later use we define a constraint-set Z_k utilizing an expression for the state predictor

$$Z_k(x_k) = \{z_k = (\pi_k, \chi_k) : \pi_k \in \Pi \ \wedge \ \chi_k \in \mathcal{X} \\ \wedge \ \chi_k = F_k(\pi_k, x_k)\} \ \forall \, x_k \in X \qquad (5)$$

3. CONVEXITY ISSUES IN OPTIMIZATION BASED CONTROL

We will in the following categorize optimization problems arising in control and consider the trade-offs which should be made. First, we separate the convex from the nonconvex cases.

3.1 *Convex problems*

We first note that problem (4) is convex if ϕ_k is a convex functional and Z_k is a (nonempty) convex set.

The issue of ϕ_k being a convex functional does not impose severe restrictions in optimizing control. The optimization criterion will typically be defined by some distance measure based on a norm, and any norm is a convex functional.

The problem of limiting $Z_k(x_k) \ \forall \ x_k \in X$ to a convex set imposes severe restrictions from a control point of view. To discuss this we first derive sufficient conditions for $Z_k(x_k)$ to be a convex set.

Sufficient conditions for convexity
The set $Z_k(x_k)$ in (5) is convex if the following assumptions are satisfied.
A1: Π is a convex set.
A2: $\mathcal{X}$ is a convex set.
A3: The mapping $F_k(\cdot, x_k) : \Pi \rightarrow R^{m_x} \times \ldots \times R^{m_x}$ is affine for all $x_k \in X$.
A4: $Z_k(x_k)$ is nonempty for all $x_k \in X$.

The proof is omitted here due to page limitations. The conditions are sufficient, since there exists nonaffine mappings F_k that renders a convex set $\mathcal{X}_{F_k}(x_k)$. This is readily seen in the 1-dimensional case where any continuous mapping provides a convex set $\mathcal{X}_{F_k}(x_k)$.

We will in the following discuss assumptions A1 to A4 above with particular emphasis on how they restrict the control problem formulation.

- Since Π is convex the set U must be convex, cf. (4). The control input at a given time instant is chosen from U, ie. $u_i \in U \subseteq R^{m_u}$. If the restrictions on the m_u different control inputs are independent U will be convex if each of the control inputs are continuous in the sense that $u_{i_j} \in [\underline{u}_j, \overline{u}_j] \subset R \ \forall j \in \{1, \ldots, m_u\}$. Defining U by including dependency between the control inputs will in most cases not introduce problems. Typical limitations like $|u_{i_j}| + |u_{i_l}| \leq 1$, $|u_{i_j} + u_{i_l}| \leq 1$, or $\|u_{i_j} + u_{i_l}\|_2 \leq 1$, where u_{i_j} and u_{i_l} denote two different control inputs at a time i, do not by themselves impose nonconvexity.
  As a conclusion Assumption A1 does not impose severe restriction on the formulation of the control problem as long as the control inputs are continuous.
- The discussion of Assumption A2 is very similar to the discussion of A1 above. Hence, if the states are continuous A2 does not impose severe restriction on the problem formulation.
- A3 is the problematic assumption since it severely limits the class of possible predictors. Viewing linear predictors, they always satisfy this assumption. Nonlinear predictors also satisfy this condition as long as they remain linear in the control input π_k.
- Assumption A4 is related to feasibility, ie. is it possible to find a control sequence Π that satisfies the state constraint $\mathcal{X}$? We will assume that this assumption always is satisfied. From a control design point of view this should be interpreted as a controllability constraint.

3.2 *Nonconvex problems*

We will now discuss nonconvex problems, and assume that Assumptions A1, A2 and A4 are satisfied. This means that nonconvexity is caused by the predictor alone. By the former discussion we claim that this is an important cause for nonconvexity, hence this type of nonconvex problems is highly relevant to optimization-based control.

3.2.1. *Identification method*

Let us first formalize an identification problem tailored for optimization-based control. The issue is to construct a predictor which produces small prediction errors on the optimization horizon, eg. I_k in (1).

First we define a set of predictors parameterized by some parameter θ, cf. (3).

$$\mathcal{F} = \{F(\cdot, \cdot; \theta) : \theta \in \Theta\}, \ \ \Theta \subset R^{m_\theta} \qquad (6)$$

For the sake of convenience and notation we have limited the set of predictors above in the sense that they are time-invariant as compared to the time-varying predictors in (3).

We assume to have a set of input and output data for the system in question and structure these in an information vector. The output is assumed to be a function of the current state. Further, we also define a truncated information vector.

$$\mathcal{I}_M = \{u_0, y_1, \ldots, u_{M-1}, y_M\} \quad (7)$$
$$y_i \in Y \subseteq R^{m_y}$$
$$y_i = g(x_i)$$
$$\mathcal{I}_i = \{u_0, y_1, \ldots, u_{i-1}, y_i\} \quad (8)$$

This allows us to define a causal predictor for the system output.

$$\hat{y}_{i+j|i} = g(\hat{x}_{i+j|i}) = g(F_j(\pi_i, \hat{x}_{i|i}; \theta)) \quad (9)$$
$$\forall i \in I_0,\ j \in I_N$$

$\hat{y}_{i+j|i}$, $\hat{x}_{i+j|i}$ – predicted output or state given $\mathcal{I}_i$

$\hat{x}_{i|i}$ – state estimate of x_i by use of $\mathcal{I}_i$

$I_0 = \{0, \ldots, M-N\}$

$j \in I_N = \{0, \ldots, N\}$

F_j refers to the j-step ahead predictor, ie. the j'th component of F_j defined in (3). Note that the predictor is time-invariant.

We observe that the prediction is based on a current state estimate $\hat{x}_{i|i}$ and the use of the predictor F_j. In the following we will make the following two simplifications. We will assume that the state is known in the sense that $\hat{x}_{i|i} = x_i$. Hence, we do not address the state-estimation problem. We further assume that the system output function g is known which means that the system output is known, ie. $\hat{y}_{i|i} = y_i$.

The identification problem is posed as

$$\theta^o = arg\ min_{\theta \in \Theta}\ V(\theta) \quad (10)$$
$$V(\theta; \mathcal{I}_M, F, g) = \sum_{i \in I_0} \sum_{j \in I_N} l_{pj}(y_{i+j}, \hat{y}_{i+j|i}) \quad (11)$$
$$l_{pj}\ :\ Y \times Y \rightarrow R$$

Let us consider the inner part of the optimality criterion. At a given time step $i \in I_0$, the inner sum computes the effect of the prediction error on the horizon $\{i+1, \ldots, i+N\}$. l_{pj} is some metric. It may vary on the prediction horizon. In particular, it can be of interest to fade away the effect of prediction errors for large values of j.

Ultimately, we would like to identify predictors which render a convex optimization problem (4). Two means of attacking this problem are (i) limiting the class of predictors $\mathcal{F}$, and (ii) modifying modify the identification criterion (11).

Limiting the class of predictors to comply with convexity leaves us with affine predictors, or even more limited, linear predictors. Viewing the identification problem the minimum criterion value $V^o = V(\theta^o; \mathcal{I}_M, F, g)$ will increase if the size of the set of predictor $\mathcal{F}$ decreases. Hence, there is a trade-off between the quality of the predictions and the advantage of having a convex optimization problem.

Referring to (ii), limiting the class of predictor can be viewed as inferring hard constraints on the optimization problem. An alternative is to include soft constraints by modifying the identification criterion in (10). In this work we do this by using ideas from (Johansen 1996) in which the identification criterion is modified to account for prior knowledge or desired properties on a problem. The modified identification criterion is given by:

$$V'(\theta\ ;\ \mathcal{I}_M, F, g) = \sum_{i \in I_0} \sum_{i \in I_N} l_{pj}(y_{i+j}, \hat{y}_{i+j|i})$$
$$+ \lambda \|S_F F\| + \gamma \|F - F_{lin}\| \quad (12)$$
$$\lambda, \gamma \in [0, \infty)$$

S_F is an operator that projects the nonlinearities of F onto a normed function space. S_F will typically contain 2nd order derivative terms which vanish if F is linear. $\|F - F_{lin}\|$ is a distance measure between an identified predictor F and the best linear predictor, F_{lin}. The norm is typically a weighted 2-norm. Let us view two possibilities. If $\gamma > 0,\ \lambda = 0$ the predictor nonlinearities are limited by a trade-off between the prediction error and the degree of nonlinearity of F. Second, if $\gamma = 0,\ \lambda > 0$ it is necessary to obtain an apriori linear predictor. The distance from this linear model is then penalized.

A problem with the above approach is to compute the norms of $S_F F$ or $F - F_{lin}$, in particular their gradients with respect to θ. It is highly desirable to have a linear parameterization since this greatly reduces the computational effort.

The proposed method may imply a significant complexity increase compared to the standard identification formulation. The identification problem is, however, an off-line problem. Further, the complexity increase is motivated by the fact that we are seeking predictors that render a convex or close-to-convex optimization-based control problem. These are on-line problems in which time-efficient algorithms can be critical for stability and performance.

The method adds soft constraints in the sense that an identified predictor may well produce a nonconvex problem. We do, however, know that a linear predictor renders a convex problem given Assumptions A1, A2 and A4. Hence, if the deviation from a linear predictor is penalized it seems reasonable to believe that this convexifies the optimization problem.

Alternative methods can be suggested by including convexity conditions as hard constraints in the identification problem.

3.3 *Convexity measure*

To the authors' knowledge there does not exist any measures of the degree of convexity in the literature. To evaluate our results in quantitative terms, however, we define a convexity measure based on the epigraph of ϕ over Π, cf. (Luenberger 1969), p.192. We define the set

$$\Phi(v; x_k) = \{\pi \in \Pi : \phi(\pi, \chi_k) \leq v \quad (13)$$
$$\wedge\ \chi_k = F_k(\pi, x_k) \in \mathcal{X}\}, \quad v \in R$$

Further we define the convex hull of Φ following Floudas (Floudas 1995), p.21.

Definition of a convexity measure
Let $\Phi(v; x_k)$ be a set (convex or nonconvex) in R^m. The convex hull, denoted $\tilde{\Phi}(v; x_k)$, of $\Phi(v; x_k)$ is the intersection of all convex sets in R^m which contains $\Phi(v; x_k)$ as a subset.

Hence, the convex hull is the smallest extension $\tilde{\Phi}(v; x_k)$ of $\Phi(v; x_k)$ such that $\tilde{\Phi}(v; x_k)$ is convex. If $\Phi(v; x_k)$ is a convex set, it is equal to its convex hull. The convexity measure is based on the volume $\Phi(v; x_k)$ compared to its convex hull.

$$\kappa = \frac{1}{\overline{v} - \underline{v}} \int_{\underline{v}}^{\overline{v}} \frac{\int_{\tilde{\Phi}(v;x_k) - \Phi(v;x_k)} d\pi}{\int_{\tilde{\Phi}(v;x_k)} d\pi} dv \quad (14)$$

By choosing $\underline{v}$ equal to the global minimum of $\phi(\pi, \chi_k)$ the denominator will always be larger than 0. $\overline{v}$ is given by

$$\overline{v} = max_{\pi_k \in \Pi}\ \phi_k(\pi_k, \chi_k) \quad (15)$$

subject to

$$\chi_k \ \in \ \mathcal{X}$$
$$h_i(x_{i+1}, x_i, u_i) = 0\ \forall\, i \in I_k$$
$$x_k \ - \ given$$

Because of the normalization the convexity measure is bounded by 0 and 1, ie. $\kappa \in [0, 1)$. The convexity measure will be computed by discretizing π uniformly, and discretizing v. The computation is extensive, in particular the calculation of the convex hull.

3.4 *Parametrization*

The identification problems posed using V' can easily become intractable. It is particularly important to choose a parametrization which limits the complexity. There are multiple choices. In this work we choose a very simple parametrization. We assume that a linear predictor and a nonlinear predictor has been identified by means of (10). A new predictor is defined as a convex combination of these two predictors.

$$F = \theta_p F_l + (1 - \theta_p) F_n, \ \ \theta_p \in [0, 1] \quad (16)$$

F_l and F_n are the linear and nonlinear predictors identified by (10). Hence, one parameter, θ_p, the weighting between the linear and nonlinear predictor, needs to be estimated. θ_p will take on values close to 0 if the influence of the terms $\lambda \|S_F F\|$ and $\gamma \|F - F_l\|$ is small compared to the prediction error term, provided F_n is a better predictor than F_l.

4. EXAMPLE

In this example we investigate properties related to a model of a continuously stirred tank reactor with the following three reactions: $A \overset{k_1}{\to} B \overset{k_2}{\to} C$, $2A \overset{k_3 x_3}{\to} D$, $E \overset{g(u)}{\to} 2E$, and state-space model formulated in continuous time.

$$\begin{aligned}
\dot{x}_1 &= -k_1(u)x_1 - k_3(u)x_3x_1^2 + \frac{v}{V}(x_{10} - x_1) \\
\dot{x}_2 &= k_1(u)x_1 - k_2(u)x_2 - \frac{v}{V}x_2 \\
\dot{x}_3 &= g(u)x_3 + \frac{v}{V}(x_{30} - x_3) \qquad (17) \\
y &= x_2 \\
k_i(u) &= k_{i0}e^{-\frac{E_i}{u}}, \ \ i \in \{1, 2, 3\} \\
g(u) &= be^{-(u-u_s)^2}
\end{aligned}$$

Parameter values are given in Table 1.

k_{10}, k_{20}, k_{30}	$1.287e+12$, $1.287e+12$, $9.043e+09$
E_1, E_2, E_3	9758.3, 9758.3, 8560.0
u_s	$387.30K$
v/V	14.18
b	10.0
x_{10}, x_{30}	5.10, 1.00

Table 1. Example 2: Parameter values.

x_1, x_2, x_3, u are the concentrations of A, B and E, and temperature, respectively, while y constitutes the process output. E is some biochemical substance which grows according to temperature conditions given by $g(u)$. It uses some inert substance to support this growth. E catalysis the reaction $2A \to D$. We assume that B is the valuable product and hence try to maximize the production of this over some time-horizon. If the reaction $E \to 2E$ is removed the model resembles a Van de Vusse reactor. We define the following criterion using a sampling time of 0.01.

$$\begin{aligned}
\phi(\pi_0, \chi_0) &= -\frac{1}{10}\sum_{i=1}^{10} |y_i| \qquad (18) \\
\pi_0 &= \{u_0, u_0, u_0, u_1, u_1, u_1, u_1, u_1, u_1, u_1\} \\
\chi_0 &= \{x_1, \ldots, x_{10}\} \\
U &= [u_s - 5.0, u_s + 5.0] \\
X &= [1.5, 2.5] \times [0.5, 2.5] \times [0.8, 2.5] \\
x_0 &- \text{given}
\end{aligned}$$

The graph of ϕ, ie. a surface in $R \times \Pi$, using the state-space model (17) for prediction, is shown in the upper left part of Fig.3 for different control inputs π and initial conditions $x_0 = (2.143, 1.090, 1.000)'$. Nonconvexity is readily observed in this case. The nonconvexity is mainly caused by $g(u)$. The global minimum for ϕ is obtained for $(u_0, u_1) = (2.15, 5.00)$. The surface changes with different values $x_0 \in X$. The nonconvexity is, however, maintained.

We will assume that the nonlinear process model is known. This forms the basis for the nonlinear predictor. This means that the optimal control input, in this case, is searched for on a graph like the upper left part of Fig.4.

A linear predictor is identified on the basis of the two solid line input data sequences shown in Fig.1 and the two corresponding output sequences shown in Fig.2. This gives 200 data points. The initial condition is $x_0 = (2.143, 1.090, 1.000)^T$. A constant level is removed from each of these data sequences. No noise is added to the data. The identification criterion is identical to (11) with l_{pj} as a quadratic function, $\hat{y}_{i|i} = y_i$ and $I_N = \{1, \ldots, 10\}$. The linear predictor is based on multiple use of the 1-step linear predictor. We use the Akaike Final Prediction Error to choose the model structure.

$$\begin{aligned} y_{i+1} &= 1.30 y_i - 0.48 y_{i-1} \\ &+ 0.0059 u_i - 0.0068 u_{i-1} \end{aligned} \tag{19}$$

Since the model is based on data in which the bias is removed the input and output in (19) must be adjusted when used for prediction. In practice the input/output data to (19) can be corrected by subtracting a lowpass-filtered value of the input and output. This is done in the prediction evaluation below.

To illustrate the prediction properties of the linear model, the 10-step ahead prediction, ie. 0.1 time units, is shown in Fig.3 for the initial condition $x = (2.010, 1.084, 1.020)'$ and a control input sequence different from the identification data set, see the validation data in Fig.1. The prediction is quite good. It should be remarked that if the control input gets close to u_s the dynamics change significantly. No identification data for control inputs in this region is available. Hence, the prediction using the linear model will deteriorate significantly for validation data in this region.

The graph of ϕ using the linear model (19) is a linear plane, thus it is convex. We use the parametrization (16) and apply (12) with $\lambda = 0$. In this case V' can be written as

$$\begin{aligned} V'(\theta_p \; ; \; \mathcal{I}_M, F, g) &= \sum_{i \in I_0} \sum_{i \in I_N} l_{pj}(y_{i+j}, \hat{y}_{i+j|i}) \\ &+ \gamma \| F_l - F_n \| (1 - \theta_p) \\ &= \sum_{i \in I_0} \sum_{i \in I_N} l_{pj}(y_{i+j}, \hat{y}_{i+j|i}) + \gamma'(1 - \theta_p) \end{aligned} \tag{20}$$

The optimal choice of θ_p increases with an increase in γ' (or equivalently an increase in γ). Further, the predictor becomes linear, ie. $\theta_p = 1$, for large values of γ'.

To study convexity we compute the convexity measure (14) for 4 different values of γ'. The measure is computed by discretizing the Π-set uniformly into 21×21 blocks, and discretizing the v-axis into 10 equidistant intervals. The results using this approximation are shown in Table 2. In this case κ is a monotonically decreasing function of θ_p. Further, the computational load and the number of failures to reach the global minimum decreases with an increase in θ_p. V_1 provides the values for the prediction error term of V' using the identification data in Figs. 1 and 2. It increases quadratically with θ_p since there is an exact match to the identification data for $\theta_p = 0$.

θ_p	0.00	0.25	0.50	0.75	1.00
V_1	0	115	460	1034	1839
κ	0.048	0.041	0.020	0.006	0.000
Average no. of functions calls	43.8	25.2	21.1	16.7	15.3
% fail to global minimum	34.5	29.9	17.0	0.2	0.0

Table 2. Example 2: The effect of different θ_p-values on the convexity measure κ, the computational load, and the search for the minimizing control inputs, π, is given.

The decision surfaces for different θ_p values are shown in Fig. 4. As can be expected the minimizing control sequence changes with varying predictors. For large θ_p values the minimizing control sequence is $\{-5.00, 5.00\}$ instead of $\{2.15, 5.00\}$ as for the correct predictor.

5. DISCUSSION

The simulation example shows that it is feasible to include convexity into the identification problem. The identification problem is affected in two ways. First, the prediction quality of the identified model decreases. This is clearly observed by viewing row V_1 in Table 2. Second, the identification problem becomes more complex. The complexity increase is, however, very modest because of the suggested parametrization (16).

The example gives a merit to the proposed method since all the tested properties of the optimization-based control problem, ie. the convexity measure,

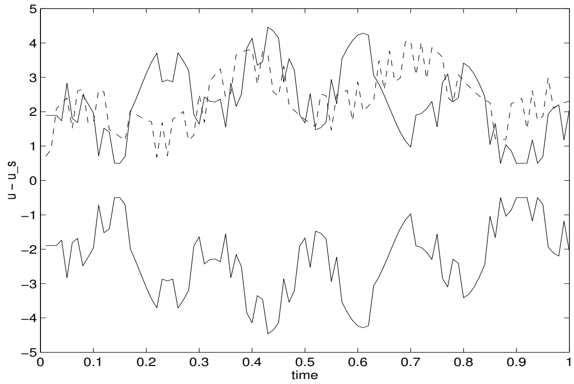

Fig. 1. Example 2: Identification input data (solid lines), validation data (broken line). Sampling time 0.01.

the computational load, and the ability to reach the global minimum, improves with an increase in γ' (or equivalently θ_p). This result can, however, not be generalized since there does exist examples where the proposed parametrization does not improve the convexity.

The results on the computational load and the failure rate to reach a global minimum will depend on the optimization method used. In this work we use a Sequential Quadratic Programming method which is the basis for the Matlab routine *constr* (Grace 1994). The motivation for this choice is the fact that this is an extensively used mainstream method.

6. CONCLUSIONS

This study shows how it is possible to take into account the convexity of an on-line optimization problem when identifying a predictor.

7. REFERENCES

Floudas, C. (1995). *Nonlinear and Mixed-Integer Optimization.* Oxford University Press.

Grace, A. (1994). *Optimization Toolbox - User's guide.* The MathWorks Inc.

Johansen, T. A. (1996). Identification of nonlinear systems using empirical data and prior knowledge - an optimization approach. *Automatica* **32**, 337–356.

Lee, J. H. (1996). Recent advances in model predictive control and other related areas. In: *Preprints CPC-V, Lake Tahoe, USA.*

Luenberger, D. G. (1969). *Optimization by Vector Space Methods.* John Wiley.

Mayne, D. Q. (1995). Optimization in model based control. In: *Preprints IFAC Symposium DYCORD, Helsingor, Denmark.*

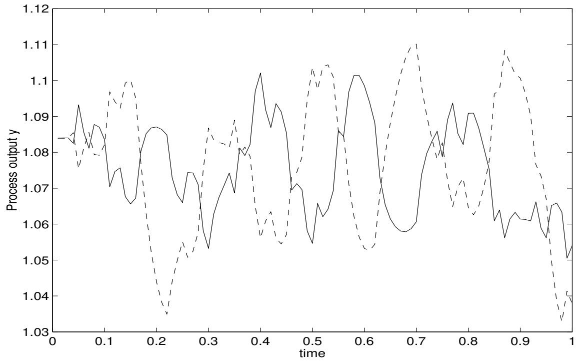

Fig. 2. Example 2: Identification process output data. Solid and broken line corresponds to input data given by the upper and lower solid lines in the previous figure. Sampling time 0.01.

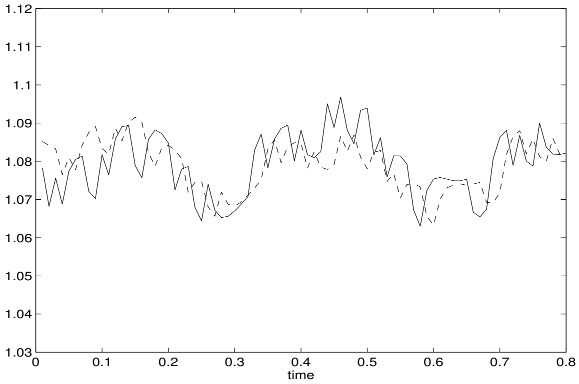

Fig. 3. Example 2: Process output data. Solid and broken line correspond to process output and 10-step ahead prediction using the linear model. Control input data is given by the broken line data in Fig.3. Sampling time 0.01.

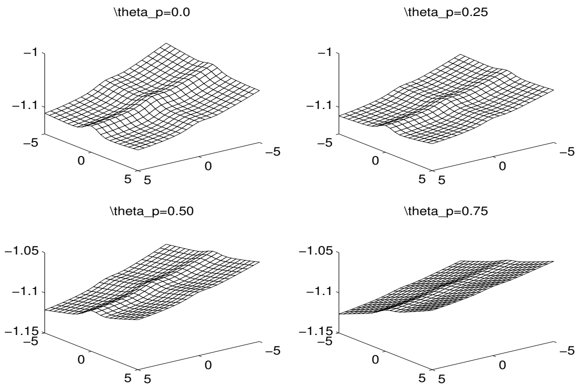

Fig. 4. Example 2: Decision surfaces, ie. $\phi(u_0, u_1)$ in (18), for $\theta_p = \{0.0, 0.250.50, 0.75\}$ are shown. $x_0 = (2.143, 1.090, 1.000)^T$.